

感情獲得プロセス理解のための MAS による 感情改善行動学習モデルの提案

茂島 祐太† 當間 愛晃‡ 赤嶺 有平‡ 山田 孝治‡ 遠藤 聡志‡

琉球大学 工学部理工学研究科† 琉球大学 工学部工学科知能情報コース‡

1. 研究背景

人間には、様々な感情がある。この感情によって人間には情動が存在し、他者とのコミュニケーションが生まれると筆者は考えている。

しかし、感情に関する様々な事柄は真に解き明かされている、とは未だ言えないのではないだろうか？感情による他者とのコミュニケーションとは決してポジティブなものだけではなく、諍い等ネガティブなものも含まれているだろう。その大半は他者の感情というものが真に理解することが未だできていないからだと考える。しかし、感情そのものを直接理解することは今までの研究からは達成することができていない。なぜならば、人間の脳を直接、あるいは間接的に観察したところで一体何を考えているのか？という対応づけがなされていないからである。

本研究では、そういった感情の真の理解のために、学習モデルを提案する。感情獲得の過程をモデル化し、観察することが出来るようになれば逆に相手の詳細な遍歴からどのような感情を抱きやすいか？ということが計算できるようになる。また、このモデルの有用性が示されれば、介護用ケアロボットのより人間らしい感情の獲得や、あるいは実際の人間の教育等で活かすこともできると考えている。

2. 用語解説

2.1. 強化学習

強化学習とは、機械学習の手法の一つであり、従来の教師あり学習や教師なし学習とは異なる位置付けにある。実際にモデルが

起こした行動を元に賞罰を決定し、その賞罰の量によって次に起こす行動を決定し学習していく。

2.1.1. A3C[1]

Deep Q-Network (DQN) の流れを汲む、強化学習アルゴリズムの一種。従来のものに比べて、非同期分散学習と、結果の先を考慮する点、従来の報酬関数(Q 関数)と併用して状態から直接行動を導き出すといった点が異なる。

特に非同期分散学習により、時系列順を破壊せずに済むため、LSTM のような時系列を考慮に入れた DNN を導入することができる。

3. 提案モデル

今回提案する学習モデルは、エージェントが行動を起こした際に周囲の反応を変化させその反応に対して報酬と状態を更新していく形式をとる。

3.1. 環境

エージェントの他に、状態を左右する外部要因が存在する。より具体的には世話人が存在することを想定し、この世話人が接触した際の行動およびエージェントの状態により報酬および次の状態が決定するものとする。

3.2. 状態

状態は、エージェントがどのような感覚を抱いているかを示す。具体的には Plutchik の提唱する感情の輪[2]を元にした基本の8つの感情、もしくは感情によらない快状態、および不快状態を持つ。

Proposal of Emotion Improvement Behavior Learning Model by MAS for Understanding Emotion Acquisition Process

† Yuta Mojima, University of Ryukyus, Faculty of Engineering, Graduate School of Science and Technology

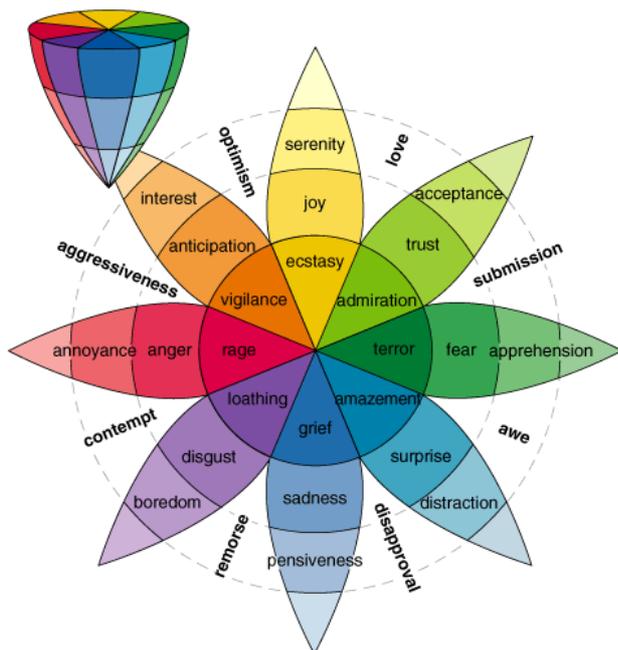


図 1. Plutchik の感情の輪

3.3. 報酬

報酬はエージェントがどれだけ快適であるかを示す。具体的にはエージェントと世話人が接触した際、世話人が行った行動が現在の状態とどれだけマッチしているかによって決定され増減する。

3.4. 行動

行動はエージェントの行動と世話人の行動がある。エージェントの行動は感情表現として想定する。世話人はそれを受けて、エージェントへの接触の可否を判定し、その後エージェントに対して行動を実施する。

4. 実験

本実験は自力行動が不可能な乳児を想定し、行動に大きく制限をかけた状態を想定した実験を行う。

4.1. 実験環境

実験にはPython3.6を使用する。

4.2. 条件設定

実験の前提条件を鑑み、以下のように各種状態を設定する。

4.2.1. 環境

主体となるエージェントである乳児および世話人一人を想定する。

4.2.2. 状態

具体的に挙げられる感情は持たず、快状態および不快状態のみを持つ。

4.2.3. 報酬

報酬は前述の通り、どれだけ快適であったかを想定する。複数ステップにまたがって快適であった場合は想定せず、世話人の行動によってのみ変化する。

具体的には、接触した際世話人がエージェントを不快状態から快状態に変化させるような行動を行なった場合は快適度を増やし、快状態から不快状態に変化させた場合は快適度を減らす。

4.2.4. 行動

エージェントが取れる行動は泣く(世話人を呼ぶ)ことだけである。

世話人はエージェントに呼ばれた場合必ず接触を行い、その際ランダムに快状態に変化させる行動、不快状態に変化させる行動、現在の状態を反転する行動のいずれかをとる。

4.3. 実験方法

モデルの条件を設定し、1000ステップの学習を行う。強化学習アルゴリズムは前述のA3Cとする。

4.4. 評価方法

実際に学習した結果が実際にどれほど現実と近いかを評価する。具体的にはエージェントが正しく不快状態の時のみ泣くようになるかを観察する。

5. 今後の展望

本実験はこれから実施予定である。今後は実験条件をより現実の人間に近いものにし、モデルが行う行動をより多種の感情表現、例えば怒ったり笑ったりするようにした時の周りの反応から、正しいタイミングで感情表現を行えるようになるかを検証する。

6. 参考文献

- [1] Asynchronous Methods for Deep Reinforcement Learning (ICML 2016)
- [2] Plutchik, R. "The Nature of Emotions" (<http://www.fractal.org/Bewustzijns-Besturings-Model/Nature-of-emotions.htm>)